OPTIMUM PLOT SIZE FOR WINTER WHEAT

by

Ronald A. Wood

Research and Development Branch
Standards and Research Division
Statistical Reporting Service

January 1972

Table of Contents

SUMMARY

Data collected from sample wheat plots in Texas, Oklahoma and Kansas in 1971 indicates that the optimum size of sample unit, in terms of minimizing sampling error, for the winter wheat objective yield survey would be between two and three rows wide, with each row somewhat less than one section long, where one section was 13 inches in length.

1. For total head weight--2.89 rows, each row .69 sections long.
2. For total number of heads--2.89 rows, each row .83 sections long.
3. For average head weight--2.59 rows, each row .64 sections long.

The optimum combination of rows and sections was assumed to be three rows, each row one section in length.

A second result of this study was that for maturity levels six and seven, there was no significant differences between units within fields for number of emerged heads. This information was based on the special wheat fields used and information drawn from regular fields in nine states. Information for nine states at maturity levels three and four showed significant differences between units for number of emerged heads. Further study should be undertaken to look at the number of units per field using a single design to measure the costs and variability at each level for analysis purposes.

OPTIMUM PLOT SIZE FOR WINTER WHEAT

by

Ronald A. Wood

INTRODUCTION

The origin of the metal frames (figure 1) now used in demarcating sample plots for the wheat objective yield surveys is not known to the author. Because of its size, it appears that it was designed to include an area of one ten thousandth of an acre.

In use, the ends of the arms of the frame are inserted consecutively through the plants at ground level parallel to a yardstick for three adjacent rows of wheat. Because of the length of the arms, they can be easily bent so they are not perpendicular to the cross piece of the frame. Consequently, the length of each row actually included in the sample plot is not necessarily the desired 26.14 inches.

This problem has long been recognized by the Data Collection Branch (DCB), Survey and Data Division, Statistical Reporting Service. They instituted an annual review of wheat frames in each state to determine which frames had become so badly bent that they were no longer usable. This procedure has not been entirely satisfactory so the DCB has recommended re-commended replacing the present wheat frame with one having shorter arms for the 1972 crop season. Since there was no information on record as to what optimum size the plot should be, the Research and Development Branch was asked to conduct a study on that subject.

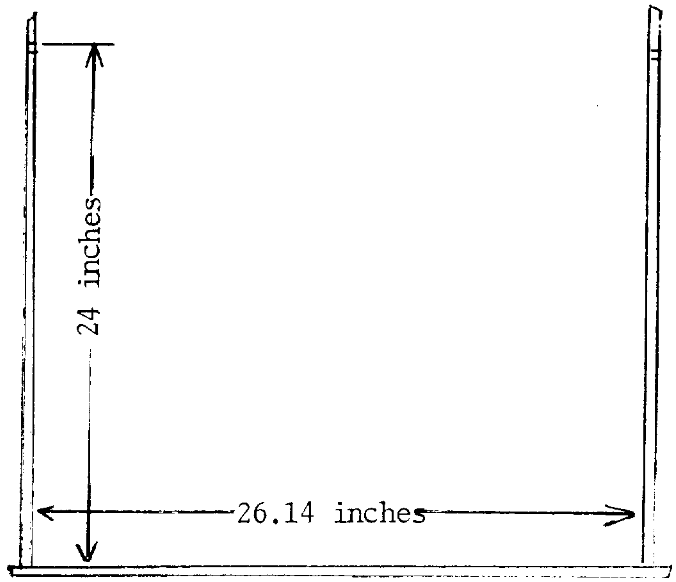


Figure 1.--Metal frame used for demarcating wheat sample plots.

DATA COLLECTION PROCEDURES

Sample Field Selection

In Kansas and Oklahoma, the regular wheat objective yield sample fields were arrayed in ascending order and a 10 percent systematic sample drawn. If an enumerator observed another crop in a sample field or found out that the wheat in the sample field was to be grazed or plowed down, the next nonspecial regular field entered by the enumerator was selected as an alternative (an alternate could be selected on first visit only). In Texas there was no alternate field provision. Texas conducted a screening survey to determine fields intended for wheat harvest. A 10 percent systematic sample was selected from the list compiled from the survey.

Field Procedures

Data for the optimum allocation study were obtained from the Unit 1 count area in the sample fields. Unit 1 was laid out for regular survey use. Work related to this project did not take place until the lowest maturity code (mc) of either Unit 1 or 2 was six. When mc 6 was reached, the rows in Unit 1 were divided in half forming six "row segments" (figure 2). The plants in each row segment were handled like the regular wheat objective yield survey. The wheat heads from each section were clipped, the total number of heads in each section recorded, and the heads were bagged by row segment and sent to the regional laboratory.

Laboratory Procedures

The laboratory work on the clipped heads was handled in the regional laboratories: Kansas in Topeka, Kansas; and Texas and Oklahoma in Oklahoma City, Oklahoma. The procedures were the same as those for the regular wheat survey, except that each row segment in Unit 1 was treated individually and then summed across to reach total weight of all heads and number of heads in sample for regular wheat objective yield work (exhibit 1).

Referring to exhibit 1 only sections 2a and 2b were used for the special survey. The weight of all heads in each row segment of Unit 1(2a1) was measured to the nearest tenth of a gram. The number of heads in each bag (2a2) was recorded in its proper row segment slot. The totals for 2a1 and 2a2 were used for the regular survey. All heads in Unit 2 were weighed together (nearest tenth of a gram) and recorded in 2b1. The total number of heads in Unit 2 was recorded in 2b2.

UNITED STATES DEPARTMENT OF AGRICULTURE
STATISTICAL REPORTING SERVICE

REGIONAL LABORATORY DETERMINATIONS

SPECIAL FORM C-2: 1971 WHEAT YIELD SURVEY – Harvested Unit Head Samples

	1– 4
Crop and Survey Number...	1 1 5 —
State (______________)..	5– 8
Segment No.	9–12
Sample No.	13–16
Row Segment	17–20
Date ______________ (Sample Processed)	21–24

1. **FROM IDENTIFICATION TAG**

	Unit 1	Unit 2	
a. Heads clipped.................... Number			29–32
b. Stage of Maturity.................. Code			33–36

2. **LABORATORY DETERMINATIONS: All Clipped Heads From Unit 1** *(by Row Segments)* **and Unit 2.**

a. Unit 1: Row Segment---	1	2	3	4	5	6	Total
(1) Total weight of all heads *(one decimal)*.... Grams	.	.	.	.	.	.	37–40 .
(2) Heads in sample..... Number							41–44

b. **Unit 2:**

(1) Total weight of all heads *(one decimal)*..............................Grams	45–48 .
(2) Heads in samples... Number	49–52

c. **Total:**

(1) Total weight of all heads 2a(1) + 2b(1).............Grams	
(2) Ten percent of 2c(1)............................Grams	
(3) Item 2c(1) minus 2c(2)..........................Grams	

3. **THRESHED GRAIN: All Heads From Units 1 and 2**

	Total
a. Weight immediately after threshing *(one decimal)*......................... Grams	53–56 .

NOTE: Is 3a less than 2c(3)?.... Yes ☐ – *Proceed to 3b.*
No ☐ – *Set forms and sample aside–NOTIFY SUPERVISOR.*

b. Weight immediately before moisture test *(one decimal)*.....................Grams	57–60
c. Moisture content 1/ .... Percent	61–64
d. Threshing loss adjustment factor .... Percent	65–68

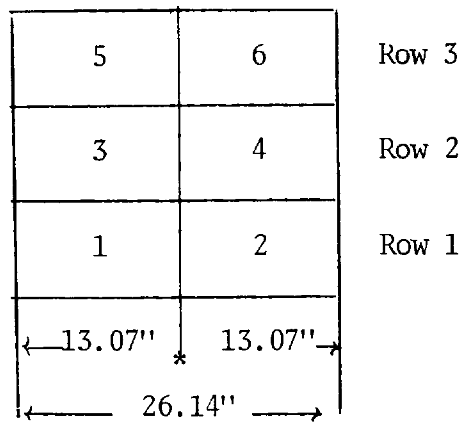
1/ *If samples are combined for moisture test, list sample numbers below. Do NOT combine sample weights in 3(a) or (b).*

____

____

HASH TOTAL | 73–80

Exhibit 1 --- Regional Laboratory Form, Special Winter Wheat Research; Objective Yield; Texas, Oklahoma and Kansas; 1971



*Measuring equipment used by enumerator could not distinguish 13.07". Therefore, some deviation from 13.07" is present.

Figure 2.--Division of regular count area into row segments for special survey: Winter wheat, Texas, Oklahoma and Kansas, 1971

ANALYSIS FOR OPTIMUM PLOT SIZE

Introduction to Optimum Procedure

The wheat objective yield sample plot has the dimensions of 26.14 inches long by three rows. "Is this the optimum plot?"

The survey design was a three stage cluster sample. The first stage of sampling was the unit (field in final analysis), the second stage was row within unit, and the third stage was section within row. The second and third stage represent an example of contiguous sampling. Contiguous sampling means that rows and sections are not randomly selected, but chosen in groups (figure 3).

A sampling design in which contiguous sampling is involved will contain a bias in the small unit (section) variance computed over the entire population. This bias is due to the selection of sections in "groups." However, the sampling design of this survey bypasses this problem through the use of variance components. This result can be attributed to the small unit variance being computed within the next higher level of sampling instead of over the entire population. Hence the optimum solution with contiguous groups contains no bias.

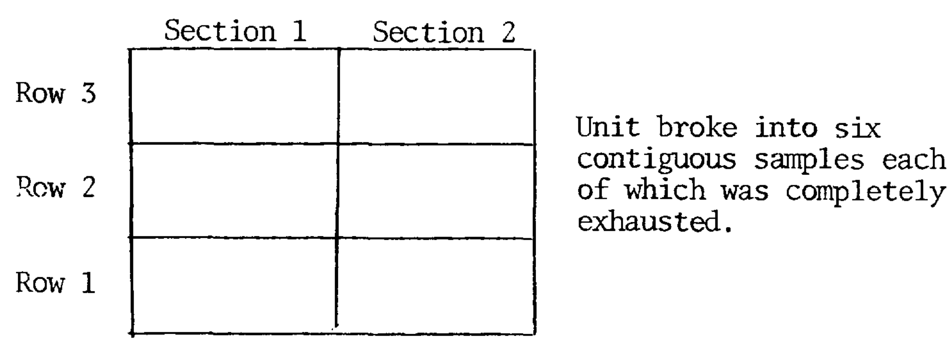


Figure 3.--Division of unit into smaller groups for sampling: Winter wheat, Texas, Oklahoma and Kansas, 1971

The optimum solution for three stage sampling is to minimize the variance x cost equation with respect to sample size n_i at each stage.

$$\text{Variance} = \frac{S_1^2}{n_1} + \frac{S_2^2}{n_1 n_2} + \frac{S_3^2}{n_1 n_2 n_3}$$

$$\text{Cost} = n_1 C_1 + n_1 n_2 C_2 + n_1 n_2 n_3 C_3$$

Where S_1^2 = variance between fields.

S_2^2 = variance between rows within units.

S_3^2 = variance between sections with rows.

C_1 = cost attributed to field.

C_2 = cost attributed to row work only.

C_3 = cost attributed to section work only.

(n_1, n_2, n_3) = sample size of respective stage sampling.

The optimum number of row sections per row (n_3) and of rows per unit (n_2) are computed as follows:

$$n_3 = \sqrt{\frac{C_2}{C_3} \frac{S_3^2}{S_2^2}} \qquad n_2 = \sqrt{\frac{C_1}{C_2} \frac{S_2^2}{S_1^2}}$$

Levels of Sampling

The regular wheat objective yield survey is hypothetically designed as a three stage sample: state, field and unit. In this study, unit one was further divided to add two additional stages of sampling, rows and sections. Since only unit one was subdivided, the levels of sampling for the optimum study are state, field, row and section. Tables 1-3 show the mean squares and F values using these four levels of sampling. The use of the unit as a level of sampling is explained in appendix II.

Table 1.--Nested analysis of variance with four levels of sampling; total weight of all heads in each section; Winter wheat, 1971

Source of variation	Degrees of freedom	Mean squares	F ratios	F .01	F .05
Between states.....	2	7295.98	7.507	4.98	3.15
Between fields.....	58	971.87	6.905	1.66	1.43
Between rows.......	122	140.75	2.245	1.32	1.22
Between sections...	183	62.71			
Totals........	365	270.90			

Table 2.--Nested analysis of variance with four levels of sampling; total number of all heads in each section; Winter wheat, 1971

Source of variation	Degrees of freedom	Mean squares	F ratios	F .01	F .05
Between states.....	2	8877.97	8.039	4.98	3.15
Between fields.....	58	1104.42	6.023	1.66	1.43
Between rows.......	122	183.36	1.871	1.32	1.22
Between sections...	183	98.02			
Totals........	365	334.57			

Table 3.--Nested analysis of variance with four levels of sampling; average weight per head in each section; Winter wheat, 1971

Source of variation	Degrees of freedom	Mean squares	F ratios	F .01	F .05
Between states.....	2	0.6610	2.654	4.98	3.15
Between fields.....	58	0.2491	8.939	1.66	1.43
Between rows.......	122	0.0279	2.477	1.32	1.22
Between sections...	183	0.0112			
Total	365	0.0581			

Cost Computation

All costs were based on approximations of time from field experience. These costs are in terms of time (table 4).

Table 4.--Approximate times to collect data from Winter wheat sample at harvest; Texas, Oklahoma, and Kansas; 1971

Stage of sampling	Average time	Time range	
		Low	High
	Minutes	Minutes	Minutes
Time associated with field (one unit)..	45.0	40.0	50.0
Time associated with one row...........	1.5	1.0	2.0
Time associated with one section.......	5.0	3.0	7.0

The time associated with one unit (field) includes time and mileage (converted to time) between fields, time to get equipment from car, time to go and return from the unit. The time associated with one row was the time to handle equipment, move between rows and write totals of all sections on ID. The time associated with one section was the time necessary to clip and count all heads and place in a paper bag.

Variance Components

The variance components for the three variables studied were computed using a nested analysis of variance. These components are shown in tables 5 through 7.

Table 5.--Variance components for total weight of heads per section; Winter wheat; Texas, Oklahoma, Kansas; 1971

Source of variation	Degrees of freedom	Mean squares	Variance components
Between states............	2	7295.98	83.76
Between fields (one unit)..	58	971.87	138.52
Between rows..............	122	140.75	39.02
Between sections..........	183	62.71	62.71

Table 6.--Variance components for total number of heads per section; Winter wheat, Texas, Oklahoma, Kansas; 1971

Source of variation	Degrees of freedom	Mean squares	Variance components
Between states............	2	8877.97	66.08
Between fields (one unit)..	58	1104.42	153.51
Between rows..............	122	183.36	42.67
Between sections..........	183	98.02	98.02

Table 7.--Variance components for average weight per head in a section; Winter wheat; Texas, Oklahoma, Kansas; 1971

Source of variation	Degrees of freedom	Mean Squares	Variance Components
Between states............	2	0.6610	0.0035
Between fields (one unit)..	58	0.2491	0.0369
Between rows..............	122	0.0279	0.0083
Between sections..........	183	0.0112	0.0112

For the purposes of the optimum study, only the variance components computed for between field (one unit), between rows, and sections were used.

S_1^2 = between fields (one unit)

S_2^2 = between rows within units

S_3^2 = between sections within rows

Optimum Number of Sampling Units - Wheat Plot

For a three stage sample design the optimum solutions for n_2 and n_3 are:

$$n_2 = \sqrt{\frac{C_1}{C_2} \frac{S_2^2}{S_1^2}} \qquad n_3 = \sqrt{\frac{C_2}{C_3} \frac{S_3^2}{S_2^2}}$$

The optimum results using total head weight per section are:

$$n_2 \ ^{1/} = \sqrt{\frac{45}{1.5} \quad \frac{39.02}{138.52}} = \sqrt{8.36} = 2.89$$

$$n_3 = \sqrt{\frac{1.5}{5} \quad \frac{62.71}{39.02}} = \sqrt{0.48} = 0.69$$

$$n_2 = 3 \text{ rows} \qquad n_3 = (0.69)$$

The optimum results using total number of heads per section are:

$$n_2 = \sqrt{\frac{45}{1.5} \quad \frac{42.67}{153.51}} = \sqrt{8.36} = 2.89$$

$$n_3 = \sqrt{\frac{1.5}{5} \quad \frac{98.02}{42.67}} = \sqrt{0.69} = 0.83$$

$$n_2 = 3 \text{ rows} \qquad n_3 = (0.83)$$

The optimum results using average head per section are:

$$n_2 = \sqrt{\frac{45}{1.5} \quad \frac{.0083}{.0369}} = \sqrt{6.72} = 2.59$$

$$n_3 = \sqrt{\frac{1.5}{5} \quad \frac{.0112}{.0083}} = \sqrt{0.405} = 0.64$$

$$n_2 = 3 \text{ rows} \qquad n_3 = (0.64)$$

1/ Optimum's for extreme times found in Appendix I.

The optimum solutions for number of rows are between two and three but closer to three. For length of section, the optimum solutions are less than one. Taking the nearest integer values the optimum solution would be for a sample plot three rows wide and one 13 inch row section long.

NONSAMPLING CONSIDERATIONS

The optimum plot size for wheat of three rows, each 13 inches in length, needs some further commentary. With a row of such short length one must consider the possibility of nonsampling errors entering into the final wheat estimates. What is known as the "edge effect" has a much greater bearing on final wheat estimates the smaller a sample plot becomes. The "edge effect" means the inclusion of extra wheat in the sample plot. The extra wheat enters the plot in two ways: (1) the arms of the frame become bent or (2) the enumerator includes borderline plants in the plot. The inclusion of this wheat introduces bias into the estimates of final wheat production. For this reason it may be logical to consider a row length at least 13 inches long.

Appendix I

An accurate time study was not practicable for this survey. Therefore, a range of possible times for each operation was defined with average times being used in the optimum solution (table 4). To support the conclusions reached in the main text, all combinations of the high and low times were used to compute optimum solutions (table 8).

Table 8.- Optimum number of rows and sections with various cost combinations: Winter wheat in Texas, Oklahoma and Kansas, 1971

Assumed costs 1/			Optimum solutions 2/	
Field	Row	Section	Rows	Sections
40	1	3	3.36	.73
40	1	7	3.36	.48
40	2	3	2.37	1.04
40	2	7	2.37	.68
50	1	3	3.75	.73
50	1	7	3.75	.48
50	2	3	2.65	1.04
50	2	7	2.65	.68

1/ All costs are in terms of time.
2/ These optimum values are based on "total head weight" per section.

The cost combination that appears most likely to cause the optimum solution to differ from that in the text is: $c_1 = 50$, $c_2 = 1$ and $c_3 = 3$. To check this, the values for the variance times cost equation were computed for various combinations of rows and sections. Variance times cost values were also calculated for average time. Both sets of results are shown in table 9.

Table 9.- Variance x cost results for total head weight per section: Winter wheat in Texas, Oklahoma and Kansas, 1971

Number of rows x number of sections	Variance times cost results using:	
	$c_1 = 45$ $c_2 = 1.5$ $c_3 = 5$	$c_1 = 50$ $c_2 = 1$ $c_3 = 3$
2 x 1	11296.98	10984.00
2 x 2	11811.21	11117.28
3 x 1	11120.84	10769.65 1/
3 x 2	12876.38	11303.00
4 x 1	nc 2/	10795.62

1/ Minimum variance x cost value.
2/ Not computed.

The optimum solution of three rows by one section holds over the range of costs.

Appendix II

The logical sequence of sampling for the wheat objective yield survey was state, field, unit, row, section. The original design only took measurements from one unit, hence eliminating the between unit variance. This made it necessary to combine two nested analyses of variance (NAV). The sources of variation in the first were rows and sections within rows. In the second NAV the sources were states, fields within states and units within fields. In the first NAV the unit of measurement was section totals. The unit of measurement in the second would have been unit totals. To place both NAV's on the same unit of measurement, the unit totals squared were divided by six to give units in terms of sections (tables 10 and 11).

Table 10.- Combined nested analysis of variance with five levels of sampling: Total weight of heads in each section, winter wheat, 1971

Source of variation	Degrees of freedom	Mean squares	Variance components
Between states....	2	17029.31	
Between fields....	58	1914.70	136.84
Between units.....	61	272.64	21.98
Between rows......	122	140.75	39.02
Between sections..	183	62.71	62.71

In computing an optimum sample design, a new variance component and cost component have been added (unit). The definitions in the main text remain the same, except for the addition of unit:

C_1 = Field 1/

C_2 = Unit - the cost to travel between units

C_3 = Row 1/

C_4 = Section 1/

Table 11.- Combined nested analysis of variance with five levels of sampling: Total number of heads in each section, winter wheat, 1971

Source of variation	Degrees of freedom	Mean squares	Variance components
Between states....	2	17683.12	
Between fields....	58	3658.58	278.44
Between units.....	61	317.69	22.38
Between rows......	122	183.36	42.67
Between sections..	183	98.02	98.02

The costs were defined in terms of time. The list below includes unit cost which was determined from field experience.

1. C_1 = 45.0 minutes
2. C_2 = 5.0 minutes
3. C_3 = 1.5 minutes
4. C_4 = 5.0 minutes

1/ Refer to cost computation section, pp. 9, 10.

The optimum solution for total weight of grain per section:

$$n_2 = \text{number of units per field} = \sqrt{\frac{C_1}{C_2}\,\frac{S_2^{\;2}}{S_1^{\;2}}}$$

$$= \sqrt{\frac{45.0}{5.0}\,\frac{21.98}{136.84}} = 1.13 \text{ units per field}$$

$$n_3 = \text{number of rows per field} = \sqrt{\frac{C_2}{C_3}\,\frac{S_3^{\;2}}{S_2^{\;2}}}$$

$$= \sqrt{\frac{5.0}{1.5}\,\frac{39.02}{21.98}} = 2.43 \text{ rows per unit}$$

$$n_4 = \text{number of sections per unit} = \sqrt{\frac{C_3}{C_4}\,\frac{S_4^{\;2}}{S_3^{\;2}}}$$

$$= \sqrt{\frac{1.5}{5.0}\,\frac{62.71}{39.02}} = .69 \text{ sections per row}$$

The optimum solution for total number of heads per section is:

$$n_2 = \sqrt{\frac{45.0}{5.0}\,\frac{22.38}{278.44}} = .72 \text{ units per field}$$

$$n_3 = \sqrt{\frac{5}{1.5}\,\frac{42.67}{22.38}} = 2.52 \text{ rows per unit}$$

$$n_4 = \sqrt{\frac{1.5}{5.0}\,\frac{98.02}{42.67}} = .83 \text{ sections per row}$$

The optimum integer solution for weight of grain is one unit by two rows by one section (table 12). Very close to this solution is one unit by three rows by one section. These two results confirm the conclusions on the plot size in the main text. The idea of one unit was not discussed in the main text, but has wide scale implications. Some further study was conducted and can be found in Appendix III.

Table 12.- Variance x cost results for total head weight per section, winter wheat, 1971

Number units x Number rows x Number sections	Variance x cost results using: $C_1 = 45$ $C_2 = 5$ $C_3 = 1.5$ $C_4 = 5$
1 x 1 x 1	14721.07
1 x 2 x 1	13210.15
1 x 2 x 2	14162.54
1 x 3 x 1	13394.73
1 x 3 x 2	16061.91
2 x 2 x 1	14034.26
2 x 2 x 2	16707.44
2 x 3 x 1	15489.78
2 x 3 x 2	19785.33

Appendix III

An interesting byproduct of this report is the result that no significant difference was observed between the units in the special survey. Since this information has large scale implications, further study was warranted. The regular survey contains information which can be used to affirm this conclusion or disprove it. In nine states, a ten percent sample was drawn of regular B-forms which were in maturity stages six and seven. The new sample design contained four stages: states, fields within states, units within fields, and rows within units. The observations used were the number of emerged heads per row on all stalks regardless of height (table 13).

Table 13.- Nested analysis of variance (NAV) computed for total number of emerged heads in each row of unit for maturity levels six and seven, winter wheat, 1971

Source of variation	Degrees of freedom	Mean squares	F ratios	$F_{.05}$	Variance components
Between states.............	8	9741.9	2.46	2.10	128.6
Between fields with state..	60	3961.2	8.32	1.47	580.9
Between units within field.	69	476.1	0.96	1.35	-6.3
Between rows within unit...	276	496.3			496.3
Total.................	413	1175.4			

The NAV shows no significant difference between units at $\alpha = .05$ and results in an indicated negative variance component. The implication is that at maturity stages six and seven only one unit is necessary for fieldwork.

The question of significant difference between units still remains for earlier maturity levels. A systematic sample was drawn for nine states from regular B-forms at maturity levels three and four. The sample design was the same used for maturity levels six and seven with observations on emerged heads used again (table 14).

Table 14.- Nested analysis of variance (NAV) computed for total number of emerged heads in each row of unit for maturity levels three and four, winter wheat, 1971

Source of variation	Degrees of freedom	Mean squares	F ratios	$F_{.05}$	Components of variance
Between states................	8	4113.6	.83	2.16	-23.98
Between fields within states..	45	4967.7	6.77	1.58	705.60
Between units within fields...	54	734.1	1.73	1.42	103.44
Between rows within unit......	216	423.8			423.83
Total....................	323	1200.15			

At this early stage of development (maturity levels 3 and 4) between field variation was very high. Conclusions about number of units within each field would have to be deferred until a cost study is made at this early level of maturity.